УДК 004.934, 621.391, 621.396.67

ВИЗНАЧЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ У МОВНОМУ СИГНАЛІ МЕТОДАМИ СПЕКТРАЛЬНО-КОРЕЛЯЦІЙНОГО ТА ВЕЙВЛЕТ-ПАКЕТНОГО ПЕРЕТВОРЕННЯ

О. О. КОРНІЄНКО, Є. А. МАЧУСЬКИЙ

Национальный технический университет Украины
«Киевский политехнический институт им. Игоря Сикорского»,
Украина, Киев, 03056, пр-т Победы 37

Анотація. Розроблено алгоритм визначення голосової активності у мовному сигналі, що грунтується на попередньому визначенні типу шумового оточення. Для опису сегментів голосних, приголосних звуків та тиші використано спектрально-кореляційний та вейвлет-пакетний методи виділення ознак мовного сигналу. Розглянуто три типи вейвлет-пакетних дерев декомпозиції, що апроксимують мел-частотну, барк-частотну шкалу та шкалу еквівалентних прямокутних смуг ERB (equivalent rectangular bandwidth) для представлення сегментів звукового сигналу. Показано, що використання двох головних компонент вейвлет-пакетних ознак дозволило з високою точністю розпізнати тип шуму оточення. Використання комбінації запропонованих ознак та структури вейвлет-пакетного дерева декомпозиції, адаптованого до критичних смуг ERB психоакустичної моделі, дозволило підвищити ймовірність правильного визначення сегментів голосу та тиші на 4% порівняно з іншими сучасними класифікаційними алгоритмами визначення голосової активності для різних типів шуму оточення.

Ключові слова: визначення голосової активності; спектрально-кореляційний аналіз; вейвлет-пакетний аналіз; критична смуга; піддіапазонні вейвлет-кепстральні коефіцієнти

1. ВСТУП

Процес визначення голосової активності у мовному сигналі полягає у відділенні мовних сегментів від сегментів тиші. Актуальною задачею є створення алгоритму визначення голосової активності, що є адаптивний до типу шумового оточення, а також до змінного співвідношення сигнал–шум SNR (signal-to-noise ratio) впродовж запису сеансу розмови [1]. Такий процес є невід'ємною складовою систем кодування мови [2], систем розпізнавання мови [3] та систем аудіо автентифікації мовця [4].

Сучасні алгоритми характеризуються зниженням ймовірності правильного виявлення голосових сегментів та сегментів тиші в умовах різного типу шуму середовища (гомін, виробничий шум, авто, тощо). Це пояснюється низкою причин, зокрема використанням: 1) простих ознак мовного сигналу [5], не здатних описати шумоподібні звуки; 2) спектральних ознак, не адаптованих для задачі визначення голосової активності [6, 7]; 3) простих порогових правил прийняття рішення [8], що не враховують нестаціонарний характер завади.

У роботі запропоновано новий алгоритм визначення голосової активності, що грунтується на використанні спектрально-кореляційного та вейвлет-пакетного методів виділення класифікаційних ознак та є вільним від зазначених недоліків.

DOI: 10.20535/S0021347018050011

БІБЛІОГРАФІЧНИЙ СПИСОК

1. Kim, Juntae; Kim, Jaeseok; Lee, Seunghyung; Park, Jinuk; Hahn, Minsoo. “Vowel based voice activity detection with LSTM recurrent neural network,” *Proc. of 8th Int. Conf. on Signal Processing Systems*, 21-24 Nov. 2016, Auckland, New Zealand. NY: ACM, 2016. DOI: 10.1145/3015166.3015207.

2. Benyassine, A.; Shlomot, E.; Su, H.-Y.; Massaloux, D.; Lamblin, C.; Petit, J.-P. “ITU-T Recommendation G.729 Annex B: a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications,” *IEEE Commun. Mag.*, Vol. 35, No. 9, P. 64-73, 1997. DOI: 10.1109/35.620527.

3. Karray, L.; Martin, A. "Towards improving speech detection robustness for speech recognition in adverse conditions," *Speech Commun.*, Vol. 40, No. 3, P. 261-276, 2003. DOI: 10.1016/S0167-6393(02)00066-3.

4. Alam, J.; Kenny, P.; Ouellet, P.; Stafylakis, T.; Dumouchel, P. "Supervised/unsupervised voice activity detectors for text-dependent speaker recognition on the RSR2015 corpus," *Proc. of Odyssey 2014: The Speaker and Language Recognition Workshop*, 16-19 June 2014, Joensuu, Finland. Joensuu, 2014, pp. 123-130.

5. Graf, S.; Herbig, T.; Buck, M.; Schmidt, G. "Features for voice activity detection: a comparative analysis," *EURASIP J. Advances Signal Processing*, Vol. 2015, P. 91, 2015. DOI: 10.1186/s13634-015-0277-z.

6. Atal, B.; Rabiner, L. "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Trans. Acoustics*, *Speech*, *Signal Process.*, Vol. 24, No. 3, P. 201-212, 1976. DOI: 10.1109/TASSP.1976.1162800.

7. Kinnunen, T.; Li, H. "An overview of text-independent speaker recognition: from features to supervectors," *Speech Commun.*, Vol. 52, No. 1, P. 12-40, 2010. DOI: 10.1016/j.specom.2009.08.009.

8. Chen, S.-H.; Wu, H.-T.; Chang, Y.; Truong, T. K. "Robust voice activity detection using perceptual wavelet-packet transform and Teager energy operator," *Pattern Recognition Lett.*, Vol. 28, No. 11, P. 1327-1332, 2007. DOI: 10.1016/j.patrec.2006.11.023.

9. Chuangsuwanich, E.; Glass, J. "Robust voice activity detector for real world applications using harmonicity and modulation frequency," *Proc. of INTERSPEECH 2011*, 28-31 Aug. 2011, Florence, Italy. ISCA, 2011, P. 2645-2648.

10. Вольфовский, Б.Н. "Многократная автокорреляционная обработка и ее возможности по обнаружению гармонического сигнала в смеси сигнала с шумом," *Информационное противодействие угрозам терроризма*, № 1, P. 91-99, 2002. URI: https://elibrary.ru/item.asp?id=9571976.

11. Madhu, S.; Bhavani, H. B.; Sumathi, S. "Performance analysis of thresholding techniques for denoising of simulated partial discharge signals corrupted by Gaussian white noise," *Proc. of Int. Conf. on Power and Advanced Control Engineering*, ICPACE, 12-14 Aug. 2015, Bangalore, India. IEEE, 2015. DOI: 10.1109/ICPACE.2015.7274980.

12. Ziolko, B.; Manandhar, S.; Wilson, R. C.; Ziolko, M. "Wavelet method of speech segmentation," *Proc. of 14th European Signal Processing Conf.*, EUSIPCO, 4-8 Sept. 2006, Florence, Italy. IEEE, 2006. URI: http://ieeexplore.ieee.org/document/7071218/.

13. Elton, R. J.; Vasuki, P.; Mohanalin, J. "Voice activity detection using fuzzy entropy and support vector machine," *Entropy*, Vol. 18, No. 8, P. 298, 2016. DOI: 10.3390/e18080298.

14. Lee, G.; Na, S. D.; Cho, J.-H.; Kim, M. N. "Voice activity detection algorithm using perceptual wavelet entropy neighbor slope," *Bio-Medical Materials and Engineering*, Vol. 24, No. 6, P. 3295-3301, 2014. DOI: 10.3233/BME-141152.

15. Rabiner, L.; Juang, B.-H. *Fundamentals of Speech Recognition*. Upper Saddle River: Prentice-Hall, 1993.

16. Fletcher, H. "Auditory patterns," *Rev. Modern Phys.*, Vol. 12, No. 1, P. 47-65, 1940. DOI: 10.1103/RevModPhys.12.47.

17. Mohammadi, M.; Zamani, B.; Nasersharif, B.; Rahmani, M.; Akbari, A. "A wavelet based speech enhancement method using noise classification and shaping," *Proc. of INTERSPEECH*, 22-26 Sept. 2008, Brisbane, Australia. ISCA, 2008, P. 561-564.

18. Sarikaya, R.; Pellom, L. Bryan; Hansen, J. H. L. "Wavelet packet transform features with application to speaker identification," *Proc. of IEEE Nordic Signal Processing Symp.*, 8-11 Jun. 1998, Vigs, Denmark. IEEE, 1998, P. 81-84. URI: https://www.isca-speech.org/archive/norsig_98/nos8_081.html.

19. Deshpande, M. S.; Holambe, R. S. "Speaker identification using admissible wavelet packet based decomposition," *Int. J. Signal Process.*, Vol. 10, No. 6, P. 83-86, 2010.

20. Добрушкін, Г.О.; Данилов, В.Я. «Порівняння якості Мел- та Барк-частотних кепстральних коефіцієнтів для параметризації мовних сигналів,» *Наукові праці Чорноморського державного університету імені Петра Могили. Сер.: Комп'ютерні технології*, Т. 160, № 148, С. 167-171, 2011. URI: http://kt.chdu.edu.ua/article/view/68900.

21. Sahu, P. K.; Biswas, Astik; Bhowmick, Anirban; Chandra, Mahesh. "Auditory ERB like admissible wavelet packet features for TIMIT phoneme recognition," *Eng. Sci. Technol. Int. J.*, Vol. 17, No. 3, P. 145-151, 2014. DOI: 10.1016/j.jestch.2014.04.004.

22. Welch, P. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, Vol. 15, No. 2, P. 70-73, 1967. DOI: 10.1109/TAU.1967.1161901.

23. Ramirez, J.; Segura, J. C.; Benitez, C.; de la Torre, A.; Rubio, A. "An effective subband OSF-based VAD with noise reduction for robust speech recognition," *IEEE Trans. Speech Audio Process.*, Vol. 13, No. 6, P. 1119-1129, 2005. DOI: 10.1109/TSA.2005.853212.

24. Thatphithakkul, N.; Kruatrachue, B.; Wutiwiwatchai, C.; Marukatat, Sanparith; Boonpiam, Vataya. "Robust speech recognition using PCA-based noise classification," *Proc. of SPECCOM*, 2004, P. 45-53.

25. Zou, Y. X.; Zheng, W. Q.; Shi, Wei; Liu, Hong. "Improved voice activity detection based on support vector machine with high separable speech feature vectors," *Proc. of 19th Int. Conf. on Digital Signal Processing*, 20-23 Aug. 2014, Hong Kong, China. IEEE, 2014. DOI: 10.1109/ICDSP.2014.6900767.

26. Garofolo, J. S.; Lamel, L. F.; Fisher, W. M.; Fiscus, J. G.; Pallett, D. S.; Dahlgren, N. L. "DARPA TIMIT

Acoustic-Phonetic Continuous Speech Corpus," NIST, 1986. URI: https://catalog.ldc.upenn.edu/ldc93s1.

27. VoxForge, Free Speech Recognition. URI: http://voxforge.org.

28. Panayotov, V.; Chen, G.; Povey, D.; Khudanpur, S. "LibriSpeech: An ASR corpus based on public domain audio books," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, ICASSP, 19-24 Apr. 2015, Brisbane, QLD, Australia. IEEE, 2015, P. 5206-5210. DOI: 10.1109/ICASSP.2015.7178964.

29. Varga, A.; Steeneken, H. J. M. "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, Vol. 12, No. 3, P. 247-253, 1993. DOI: 10.1016/0167-6393(93)90095-3.

30. Корнієнко, О.О. "Вейвлет-пакетні ознаки мовного сигналу у завданні розпізнавання мовця," *Вимірювальна та обчислювальна техніка в технологічних процесах*, № 2, С. 111-117, 2017.

31. Корнієнко, О.О.; Кущ, С.М. "Адаптивний алгоритм визначення голосової активності," *Матеріали конференції «Радіотехнічні поля, сигнали, апарати та системи»*. URI: http://conf.rtf.kpi.ua/attachments/article/490/RTPSAS_2015_s8_t04.pdf.

32. Friedman, J. H. "Another Approach to Polychotomous Classification," Technical Report. Department of Statistics, Stanford University, 1996, P. 1-14. URI: http://www-stat.stanford.edu/~jhf/ftp/poly.ps.Z.

33. Chang, C.-C.; Lin, C.-J. "LIBSVM: A library for support vector machines," *ACM Trans. Intelligent Syst. Technol.*, Vol. 2, No. 3, Article No. 27, 2011. DOI: 10.1145/1961189.1961199.

34. Ramyrez, J.; Yélamos, P.; Górriz, J. M.; Segura, J. C.; García, L. "Speech/non-speech discrimination combining advanced feature extraction and SVM learning," *Proc. of 9th Int. Conf. on Spoken Language Processing*, 17-21 Sept. 2006, Pittsburgh, Pennsylvania. 2006, P. 1662-1665.

35. Zhang, Y.; Tang, Z.-M.; Li, Y.-P.; Luo, Y. "A hierarchical framework approach for voice activity detection and speech enhancement," *The Scientific World Journal*, Vol. 2014, Article ID 723643, 2014. DOI: 10.1155/2014/723643.

36. Sohn, J.; Kim, N. S.; Sung, W. "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, Vol. 6, No. 1, P. 1-3, 1999. DOI: 10.1109/97.736233.

37. Eyben, F.; Weninger, F.; Squartini, S.; Schuller, B. "Real-life voice activity detection with LSTM recurrent neural networks and an application to Hollywood movies," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, ICASSP, 26-31 May 2013, Vancouver, BC, Canada. IEEE, 2013, P. 483-487. DOI: 10.1109/ICASSP.2013.6637694.

Поступила в редакцию 15.03.2017 После переработки 06.03.2018